Investigating the influence of special on-off attacks on challenge-based collaborative intrusion detection networks

Li, Wenjuan; Meng, Weizhi; Kwok, Lam For

Published in:
Future Internet

Link to article, DOI:
10.3390/fi10010006

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

future internet

MDPI

Article

Investigating the Influence of Special On–Off Attacks on Challenge-Based Collaborative Intrusion Detection Networks †

Wenjuan Li [1], Weizhi Meng [2,*] and Lam For Kwok [1]

[1] Department of Computer Science, City University of Hong Kong, Hong Kong, China; wenjuan.li@my.cityu.edu.hk (W.L.); cslfkwok@cityu.edu.hk (L.F.K.)

[2] Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

* Correspondence: weme@dtu.dk; Tel.: +45-4525-3068

† A preliminary version of this paper has been presented at the 12th International Conference on Green, Pervasive, and Cloud Computing (GPC), 2017; pp. 402–415.

Received: 15 December 2017; Accepted: 3 January 2018; Published: 8 January 2018

Abstract: Intrusions are becoming more complicated with the recent development of adversarial techniques. To boost the detection accuracy of a separate intrusion detector, the collaborative intrusion detection network (CIDN) has thus been developed by allowing intrusion detection system (IDS) nodes to exchange data with each other. Insider attacks are a great threat for such types of collaborative networks, where an attacker has the authorized access within the network. In literature, a challenge-based trust mechanism is effective at identifying malicious nodes by sending challenges. However, such mechanisms are heavily dependent on two assumptions, which would cause CIDNs to be vulnerable to advanced insider attacks in practice. In this work, we investigate the influence of advanced on–off attacks on challenge-based CIDNs, which can respond truthfully to one IDS node but behave maliciously to another IDS node. To evaluate the attack performance, we have conducted two experiments under a simulated and a real CIDN environment. The obtained results demonstrate that our designed attack is able to compromise the robustness of challenge-based CIDNs in practice; that is, some malicious nodes can behave untruthfully without a timely detection.

Keywords: intrusion detection; collaborative network; on–off attack; challenge-based mechanism; trust computation and management

1. Introduction

The major goal of an intrusion detection system (IDS) is to identify any signs of suspicious activities in either systems or networks [1]. IDSs are widely adopted in various organizations and can be generally classified into two groups: *host-based IDSs (HIDSs)* and *network-based IDSs (NIDSs)*. The HIDS identifies malicious events for an end system or application by monitoring local events and states. The NIDS focuses on network environments and detects potential attacks by monitoring and examining traffic outside the demilitarized zone (DMZ) or within an internal network [2]. Further, there are usually two major detection methods for a typical IDS, namely, the signature-based detection approach and the anomaly-based detection approach.

A signature-based IDS detects suspicious events by comparing incoming payloads with stored signatures (called rules), while an anomaly-based IDS detects malicious events through identifying significant deviations between the current behavioral profile and the normal behavioral profile. A normal profile is used to describe the characteristics of applications and connections via monitoring for a period of time [1]. With the increasing complexity of current intrusions, it is found that a single

or isolated IDS would not work effectively in a complicated scenario [3,4]. These attacks may cause great damage if they cannot be detected timely; that is, they may cause the entire network to be paralyzed. With the purpose of improving the detection accuracy of single IDSs, research has been made for collaborative intrusion detection networks (CIDNs), which enable different IDS nodes to collect and exchange data with each other [4]. The collaborative nature of CIDNs can help to optimize the capability of an IDS; however, insider attacks are one great threat that can significantly degrade the security level of the whole network [3]. As a result, there is a need to implement additional mechanisms to protect a collaborative environment itself.

Building appropriate trust-based mechanisms is a promising solution to protect CIDNs against insider threats. For this purpose, Fung et al. [5] developed a kind of challenge-based trust mechanism (or *challenge mechanism*) for CIDNs, which utilizes *challenges* to evaluate the reputation of IDS nodes. A challenge may contain some predefined alarms requesting the target node to rank the severity. Because the testing node generates the challenge (i.e., extracting from its database), it knows the alarm severity in advance. The reputation of an IDS node can be judged according to the satisfaction level between the expected answer and the received feedback. A line of relevant studies (e.g., [5–7]) have proven that the challenge mechanism can be robust against common insider attacks, like collusion attacks, in which several adversarial nodes work together to provide fake alarm information to target nodes, aiming to degrade the detection effectiveness.

Motivations. The challenge mechanism has shown good performance against common insider attacks, but it depends heavily on two major assumptions: (1) it is hard for an IDS node to distinguish between a challenge and a normal message; (2) malicious nodes would always send untruthful feedback. In a practical implementation, these two assumptions are not realistic in most cases, as adversarial nodes can behave in a much more dynamic and complicated way [8,9]. As a result, because of these assumptions, challenge mechanisms may become problematic under some advanced attacks. As an example, Li et al. [8] designed an advanced attack, named the *passive message fingerprint attack* (PMFA), which could help to distinguish between a challenge and normal messages. Under the PMFA, an IDS node can send untruthful answers to normal requests without decreasing their trust values.

Contributions. In this work, our motivation is to investigate the influence of a special on–off attack (SOOA), which is able to behave normally to one node while sending untruthful answers to another node. Differently from the previous version [10], this work further evaluates the attack performance of the SOOA in a real network environment. The contributions of this work are listed below:

- We first describe the high-level architecture of a typical challenge-based CIDN with the adopted assumptions and then investigate the influence of the SOOA, which can behave normally to one IDS node while responding maliciously to another node. In this case, trust computation in the third node may be affected, as it may receive the opposite judgement from its partner nodes.
- To investigate the attack performance, we have performed two experiments under a simulated and a real CIDN environment. Our results demonstrate that the SOOA has the potential to greatly affect the trust computation of IDS nodes; that is, some malicious nodes can keep their reputation without timely detection. Finally, we discuss some countermeasures and solutions.

Different from the previous work [10], this work both further evaluates attack scenarios and has performed an evaluation in a real CIDN environment. We acknowledge that challenge mechanisms are a promising solution to safeguard CIDNs against malicious insider nodes. The purpose of our work is to attract more research efforts to enhance the application of challenge mechanisms in practical scenarios.

The remaining parts are organized as follows. Section 2 presents a set of related work regarding trust management in distributed IDS networks. Section 3 introduces the architecture of challenge-based CIDNs and analyzes the adopted assumptions. Section 4 describes how (SOOA) works in a challenge-based CIDN and discusses two scenarios as a study. Section 5 describes two major experiments under a simulated and a real CIDN environment. Finally, Section 6 concludes our work with future directions.

2. Related Work

Collaborative intrusion detection systems/networks are developed to boost the accuracy of a separate detector, which usually has less information about the protected environment. This collaborative network enables various IDS nodes to request and collect data from other nodes. However, the collaborative nature renders it vulnerable to insider attacks, in which intruders are inside the network. To protect distributed systems and collaborative networks against malicious nodes, establishing a proper trust-based intrusion detection mechanism is desirable.

Trust-Aware Mechanism

Trust management has been widely studied in literature. Duma et al. [3] described a P2P-based overlay IDS, which utilizes a trust engine to handle alarms and an adaptive scheme to calculate reputation. More specifically, the former is used to filter out alerts sent by untrusted or low-reputation nodes, while the latter can calculate the reputation of nodes by considering their past behaviors. Meng et al. [11] recently proposed a Bayesian inference-based trust mechanism to identify untruthful nodes for medical smartphone networks. The evaluation showed that their approach could quickly identify malicious nodes in real scenarios. For some other related works, we refer to [12–18].

Challenge-Based Trust Mechanism

How to design an appropriate trust management in CIDNs remains an issue. For this purpose, Fung et al. [5] designed a challenge-based trust mechanism, which sends challenges to evaluate the reputation of an IDS node. The trustworthiness of a node can be derived according to the received answers. At first, they described a detection framework based on HIDSs, in which each HIDS node could judge the trustworthiness of others on the basis of the difference between the sent challenges and the received answers. They further utilized a forgetting factor to emphasize the recent feedback [6]. Then, they enhanced their mechanism with a Dirichlet-based model, which allows for the evaluation of the reputation of IDS nodes by considering their mutual behavioral events [7]. In the evaluation, they mainly evaluated their model for challenge-based CIDNs in some simulated environments. The mechanism was found to have strong scalability properties and to be robust against common insider threats.

Advanced Insider Attack

Current intrusions have become more complex, and many research studies have moved to advanced attacks. Li et al. [8,19] developed an advanced collusion attack, named *passive message fingerprint attack* (PMFA), which allows several malicious nodes to exchange received data and distinguish normal requests passively. Experimental results indicated that the PMFA enabled IDS nodes to give untruthful answers to normal requests without decreasing their trust values. Similarly, Meng et al. [9] also developed an advanced collusion attack, called the *random poisoning attack*, which enables a node to provide malicious answers with a predefined possibility. They performed two experiments under both simulated and real environments, and it was found that this attack could compromise the robustness of challenge-based CIDNs.

Mechanism Improvement

To enhance the mechanism performance, Li et al. [20] pointed out that distinct IDS nodes may not have the same detection capabilities. Some nodes could have a higher or lower level of sensitivity for the detection of some particular intrusions. As an example, the number of signatures can decide whether an IDS node has a stronger capability of identifying a kind of virus. That is, a node can be more accurate in identifying such a threat if it has a larger set of relevant signatures. On the basis of this observation, they proposed *intrusion sensitivity* (IS), which could be used to measure the detection sensitivity of an IDS node in terms of particular intrusions. They further proposed a trust management

approach by means of IS, through automating the allocation of IS with machine learning techniques in real-world applications [21,22]. Pollution attacks are a kind of insider threat that allow a set of malicious nodes to work collaboratively to give fake alarm information to the target node. Li and Meng [23] conducted a study to explore the influence of IS on the detection of pollution attacks. It was found that this notion can help to detect malicious nodes quickly by emphasizing the impact of expert nodes.

3. Challenge-Based CIDNs

3.1. Background

To protect collaborative networks against insider attacks, many trust-based approaches have been proposed [24]. Challenge mechanisms are one effective approach to point out unusual nodes and measure the trustworthiness of nodes according to the received feedback [5]. Figure 1 presents a typical challenge-based CIDN with major components of an IDS node.

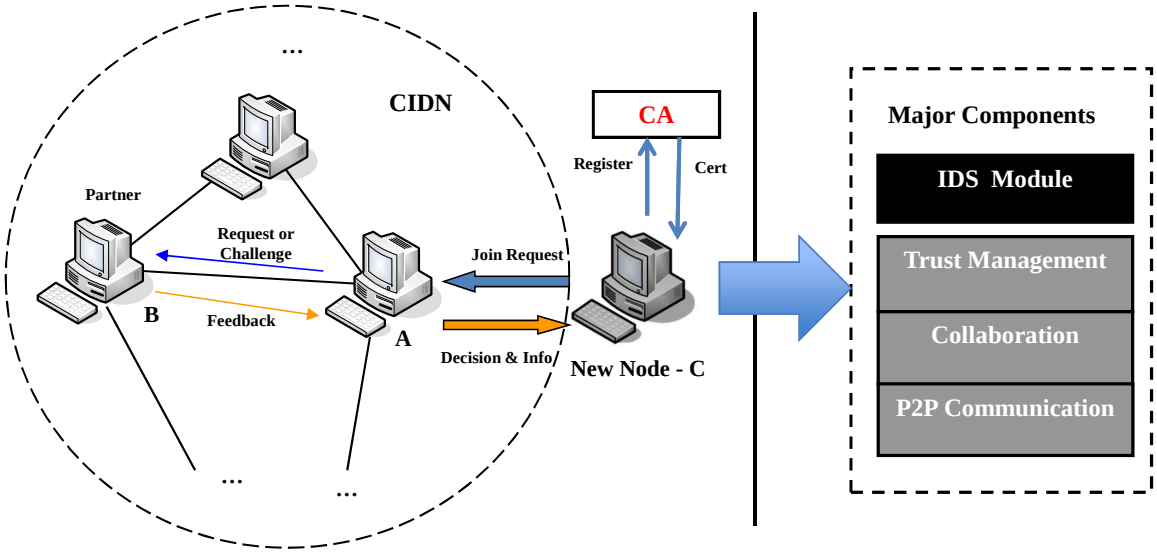

Figure 1. The high-level architecture of a typical challenge-based collaborative intrusion detection network (CIDN).

In such networks, IDS nodes can choose their own collaborators or partners in terms of their prior experience, as well as maintain a list of connected partners. This list is known as a *partner list* (or *acquaintance list*), and it can gather necessary information with other IDS nodes, for example, public keys and reputation levels. Supposing an outside node plans to join the network, it has to firstly obtain its proof of identify by registering via a trusted certificate authority (CA), including a public and private key pair. As shown in Figure 1, if node *C* plans to join the CIDN, it can apply to a node within the network, for example, node *A*. After receiving a request, node *A* can make decisions on the basis of the predefined rules and return a list of initial partners if the request is confirmed.

Interactions

To improve the detection accuracy of a separate IDS node, collaborative networks enable many IDS nodes to exchange data with other nodes; that is, several nodes can exchange alarm information to obtain a high-level view of the network status. In a challenge-based CIDN, two types of messages would be used during node interactions.

- *Challenges.* This type of message contains several IDS alarms requesting the target node to rank the severity. For instance, a testing node can send a challenge periodically to one or several tested

nodes and then obtain their answers. Because the testing node extracts IDS alarms from its own database, it can know the alarm severity in advance. Accordingly, it can evaluate the tested nodes' trustworthiness by identifying the deviation between the expected and the received feedback. For the satisfaction mapping, we refer to Section 5.

- *Normal requests.* This type of message is sent by a detector to collect data for alarm aggregation. In a CIDN, if a node starts to aggregate alarms, it can send a normal request to other IDS nodes. Then, other trusted nodes can give alarm information on the basis of their own experience. Intuitively, alarm aggregation is a very important step to improve the detection accuracy of a separate intrusion detector. It is worth noting that the alarm aggregation process only considers the information from trusted nodes.

Major Components

As shown in Figure 1, an IDS node contains an *IDS module* and consists of three major components, the *trust management component*, *collaboration component* and *P2P communication*.

- *Trust management component.* To measure the reputation of IDS nodes, this component is responsible for comparing the expected answer with the received feedback. As mentioned above, each IDS node can request for the alarm severity through sending either normal requests or challenges. In order to protect challenges, Fung et al. [5] assumed that challenges should be delivered randomly, associated with timing, making them hard to be identified from a normal request.
- *Collaboration component.* The goal of this component is to handle CIDN messages, that is, to help a node measure the reputation of others by sending normal requests or challenges. For example, this component can return the answers when an IDS node receives a CIDN message. In Figure 1, node *B* would return its feedback according to its own experience, if node *A* delivers a request or a challenge.
- *P2P communication.* This component aims to help maintain connections with other nodes, that is, by configuring the network initialization, address management and node-to-node communication. The trust of the P2P communication is assumed to be trusted.

Robustness

A line of research studies (e.g., [5–7]) have shown that challenge-based trust mechanisms can protect CIDNs against common threats such as a sybil attack, a newcomer (re-entry) attack, and a betrayal attack.

- *Sybil attack.* This kind of attack describes the situation in which a node tries to create many fake identities [25]. These fake identities can be utilized to gain a larger impact on alarm aggregation in a CIDN node. The challenge mechanism mitigates this attack through requesting each IDS node to register via a trusted CA and obtain a unique proof identity.
- *Newcomer (re-entry) attack.* This type of attack indicates the situation in which a node tries to re-enter the network as a newcomer, aiming to erase its bad history. The challenge mechanism avoids this attack by allocating a low reputation level to all new joined nodes.
- *Betrayal attack.* This kind of attack indicates the situation in which a trusted node turns out to be an untruthful node unexpectedly. The challenge mechanism mitigates this attack by employing a strategy: that is, a high reputation level can only be achieved after a long time-period of interaction with consistent good behavior, whereas the reputation can be quickly degraded by detecting only a few bad actions. To realize this strategy, a forgetting factor can be used to give more weight to recent behavioral events.

Overall, challenge-based CIDNs can encourage collaborations among various IDS nodes, as well as identify common insider attacks. However, it is found that challenge-based CIDNs would suffer

from advanced insider attacks, because the adopted assumptions are not realistic in real-world implementations [8,20,21].

3.2. Assumption Analysis

Challenge-based CIDNs are shown to be robust against several common insider attacks in prior studies [6,7]. However, challenge mechanisms rely on two major assumptions, causing CIDNs to be problematic in real implementations.

- *First assumption.* Challenges should be sent randomly to ensure they are hard to be identified from normal messages.
- *Second assumption.* Malicious nodes always behave untruthfully to other nodes, that is, by sending untruthful answers to the received messages.

The first assumption indicates that an IDS node cannot distinguish a challenge from normal messages, ensuring that it has a small possibility for a malicious node to give manipulated feedback to challenges. However, this assumption still leaves a chance for attackers to figure out the challenges in practice. In literature, Li et al. [8] designed an advanced attack, which can distinguish normal requests from messages with a high probability and enable nodes to send untruthful feedback only to normal requests without decreasing their trust levels.

The second assumption attempts to ensure that malicious nodes always behave abnormally, which helps to decrease their reputation in a fast manner. Fung et al. [5] summarized this assumption as a *maximal harm model*, in which a malicious node chooses to give untruthful answers with the purpose of making the most negative influence on the target nodes. However, in a practical scenario, a malicious node can choose a different harm model and send malicious answers in a dynamic way to keep the trust level.

Discussion. The above assumptions are realistic in some scenarios in which an intruder is naive and willing to use the maximal harm model. However, many intruders may choose a different and dynamic strategy to attack CIDN nodes. For instance, a malicious insider node can give untruthful feedback to some nodes while behaving normally to other nodes. On the whole, the assumptions adopted by existing challenge mechanisms are too strong for real-world implementations, leaving a chance for advanced attackers to compromise the CIDN security.

4. SOOA: Special On–Off Attack

On the basis of the above analysis, challenge-based CIDNs may suffer from advanced attacks in real scenarios. In this work, our motivation is to explore the impact of a SOOA on the robustness of challenge mechanisms. Our attack allows a node to respond truthfully to certain nodes but behave maliciously to the other nodes. It is worth emphasizing that we only accept the first assumption but improve the second assumption: that is, a malicious node can deliver untruthful answers with a strategy.

On–Off Attacks

In literature, a general on–off attack indicates the situation in which an attacker behaves well and badly alternately, aiming to compromise the network if they remain as trusted nodes [26]. The type of attack has two major states: *on-state*, when the associated action is effectively happening, and *off-state*, when the associated action is not happening. By behaving as a good node and as a bad node alternately, this type of attack pretends to be a temporary error for a security mechanism. A balance is often made: that is, a high ratio of off-state in relation to on-state is a more effective attack, while a low ratio might make it more easy for a trust management scheme to detect the malicious behavior.

Special On–Off Attack

As a study, this work considers a SOOA, for which a malicious node can keep sending truthful answers to one node but behave maliciously to another node. This particular attack has the potential to affect the effectiveness of the trust computation for a third node (target node). This attack accepts that a challenge could be delivered randomly to make it difficult to be identified from normal messages. The sending randomness can be achieved by a random number generator. Figure 2 describes an example of the SOOA: supposing node *D* is malicious, node *A* is the target, and node *B* and node *C* are partner nodes of node *A*. Two adversary scenarios are described below.

- **Scenario 1: node *D* is not a partner node of node *A*.** In this scenario, node *D* chooses to send a truthful response to node *C* while sending untruthful (or malicious) answers to node *B*. Figure 2 shows that node *A* can communicate and collect data with/from its partner nodes; thus, node *A* may receive different (or opposite) reports on node *D*. This scenario often occurs for a hierarchical network structure, for which a central server needs to collect information and judge the trustworthiness of each node.
- **Scenario 2: node *D* is a partner node of node *A*.** In this scenario, node *D* can respond truthfully to node *A* if they are partner nodes. In CIDNs, node *A* can judge the trustworthiness of node *D* through both its own trust computation and the judgement from other nodes, for example, nodes *B* and *C*. In this case, node *D* can perform the same as in Scenario 1 to affect the decision, such as by alarm aggregation of node *A*.

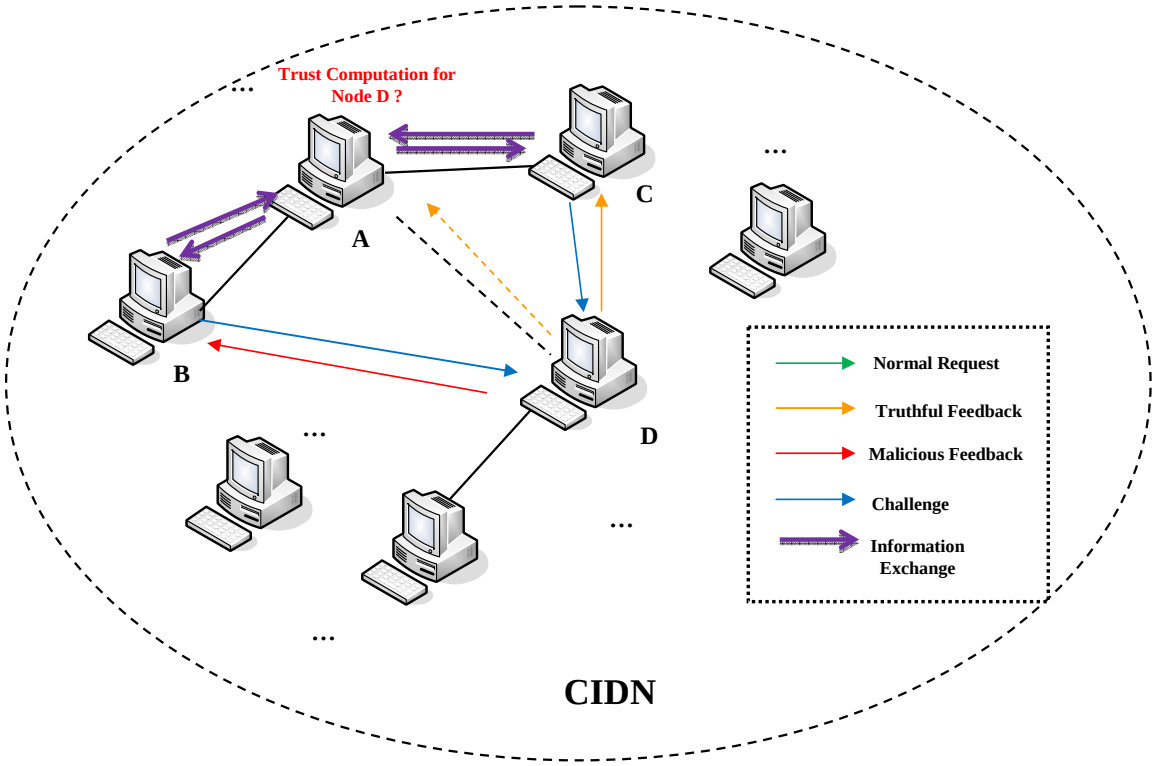

Figure 2. An example of a special on–off attack (SOOA) on challenge-based collaborative intrusion detection networks (CIDNs).

On the whole, our SOOA can choose to give truthful feedback to several nodes while responding untruthfully to others. As a result, it may affect the trust computation of certain nodes and maintain its trust values above the threshold, that is, maintaining the trust values of node *D* above the threshold

as for node A. In practice, malicious nodes can make a negative impact on alarm aggregation in both scenarios, through maintaining their trust values.

5. Evaluation

In this section, we measure the performance of the SOOA in both a simulated and a real CIDN environment, under either Scenario 1 or Scenario 2, respectively. In the remaining parts, we describe the deployment of CIDNs and the calculation of reputation levels, as well as discuss the obtained results.

5.1. CIDN Settings

In the simulated CIDN environment, a total of 15 nodes were distributed randomly within a 5×5 grid region. Each node employs an open-source IDS plugin (Snort [27]) and builds a partner list after connecting to other IDS nodes. Similarly to a previous work [7], we chose the initial trust levels of all nodes to be $T_s = 0.5$.

To measure the reputation levels of partner nodes, each node can deliver challenges in a random manner with an average rate of ε. Two request frequencies are considered in this work: ε_l and ε_h. The request frequency would be at a low level for highly trusted or highly untrusted nodes, because their reputation levels are relatively stable. In contrast, the request frequency would be at a high level for the nodes whose trust levels are close to the detection threshold. In this work, we selected the low request frequency to be 10 per day, which was more strict than for the previous work (e.g., [7]). Table 1 shows the detailed parameters for the deployed CIDN.

Table 1. Simulation parameters in the experiment.

Parameters	Value	Description
λ	0.9	Forgetting factor
ε_l	10/day	Low request frequency
ε_h	20/day	High request frequency
r	0.8	Trust threshold
T_s	0.5	Trust value for newcomers
m	10	Lower limit of received feedback
d	0.3	Severity of punishment

Node Expertise

Similarly to former studies, this work also considered three expertise levels for a CIDN node: low (0.1), medium (0.5) and high (0.95). The IDS's expertise can be modeled by means of a beta function as shown below:

$$\begin{aligned} f(p'|\alpha,\beta) &= \frac{1}{B(\alpha,\beta)} p'^{\alpha-1}(1-p')^{\beta-1} \\ B(\alpha,\beta) &= \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \end{aligned} \tag{1}$$

where $p'(\in [0,1])$ describes the probability of an IDS in examining an intrusion, and $f(p'|\alpha,\beta)$ describes the probability that an IDS node at the expertise level l responds in p' to an intrusion examination of difficulty level $d(\in [0,1])$. A higher l value means a larger probability of correctly detecting an intrusion, while a higher d value means that an intrusion is harder to identify. Further, α and β can be set as below:

$$\begin{aligned} \alpha &= 1 + \frac{l(1-d)}{d(1-l)} r \\ \beta &= 1 + \frac{l(1-d)}{d(1-l)} (1-r) \end{aligned} \tag{2}$$

where $r \in \{0,1\}$ is the expected result of the detection. Regarding a fixed difficulty level $d(\in [0,1])$, the node with higher expertise should have a larger probability of correctly identifying an intrusion. For instance, a node with an expertise level of 1 can accurately identify an intrusion if the difficulty level is 0.

Node Trust Evaluation

An IDS node can send a challenge periodically to measure the reputation of a tested node. A satisfaction level can be derived on the basis of the difference between the expected answers and the received feedback. As a result, the trustworthiness of a node i according to node j can be computed as below:

$$T_i^j = (w_s \frac{\sum_{k=0}^{n} F_k^{j,i} \lambda^{tk}}{\sum_{k=0}^{n} \lambda^{tk}} - T_s)(1-x)^d + T_s \quad (3)$$

where $F_k^{j,i} \in [0,1]$ represents the satisfaction level of a received feedback k, n is the total number of received feedbacks, λ is the *forgetting factor* by allocating more weight to recent answers, and w_s is the *significant weight*, which relies on the number of received feedbacks. If the number of received feedbacks is below a minimum threshold m, then $w_s = \frac{\sum_{k=0}^{n} \lambda^{tk}}{m}$; otherwise $w_s = 1$; x is the percentage of "don't know" replies for a period of time; d is a positive incentive parameter, which is used to control the severity of the punishment to "don't know" replies. For the detailed equation derivation, we refer to [6,7].

Satisfaction Evaluation

We let $(e \in [0,1])$ denote an expected feedback and $(r \in [0,1])$ denote a true received feedback. A function F $(\in [0,1])$ can be defined to calculate the satisfaction level by identifying the difference between the the expected feedback and the received feedback [7]:

$$F = 1 - (\frac{e-r}{max(c_1 e, 1-e)})^{c_2} \quad e > r \quad (4)$$

$$F = 1 - (\frac{c_1(r-e)}{max(c_1 e, 1-e)})^{c_2} \quad e \leq r \quad (5)$$

where c_1 and c_2 control the degree of penalty for wrong estimates and the satisfaction sensitivity, respectively. Similarly to a former study [7], this work sets $c_1 = 1.5$ and $c_2 = 1$ in the evaluation.

5.2. Simulation Experiment

In this experiment, we aimed to investigate the influence of the SOOA on challenge-based CIDNs under both Scenario 1 and Scenario 2. Following the example given in Figure 2, we set up two scenarios as below:

- **Scenario 1.** In this condition, node A had six partner nodes in its list without node D. In this case, node D behaved normally to some partner nodes of node A, while it behaved untruthfully to the remaining partner nodes.
- **Scenario 2.** In this scenario, node A had seven partner nodes including node D, which could send truthful answers to several partner nodes of node A, while sending untruthful (or malicious) answers to the rest of the partner nodes.

Figure 3 depicts the convergence of trust values for different expert nodes: low ($I = 0.1$), medium ($I = 0.5$) and high ($I = 0.95$). The results validated the observations obtained in previous studies [6,7]: that is, nodes with higher expertise can achieve bigger trust values. The trust values of all nodes became stable after around 16 days in the simulated network.

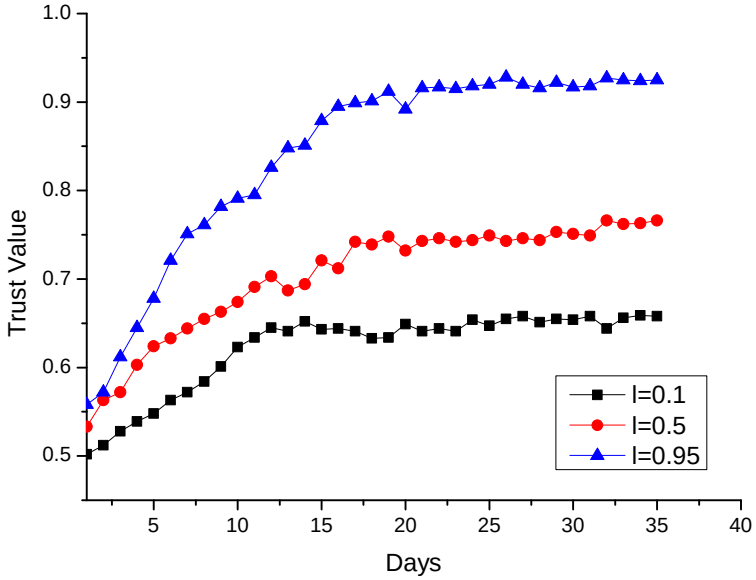

Figure 3. Convergence of trust values of intrusion detection system (IDS) nodes regarding three expertise levels.

Results under Scenario 1

In this condition, we randomly selected one expert node ($I = 0.95$) to perform our attack from day 45. As node D was not a partner of node A, node A could only evaluate the trustworthiness of node D by collecting the data from its partner list. As a study, we tested two conditions: (1) node D sent truthful answers to four partner nodes of node A, but behaved untruthfully to the other two partner nodes (T4U2); (2) node D responded truthfully to three partner nodes of node A, while sending untruthful answers to the remaining partner nodes (T3U3). Figure 4 describes the trust value of the malicious node.

- *T4U2*. Under this condition, it was found that the trust value of the malicious node (node D) could gradually decrease below the threshold after nearly 15 days, while the trust value could return above the threshold over a period of time. This was because up to four partner nodes reported a "benign" status for node D. Supposing node A was a central server in a CIDN, node D could still make an impact on its judgement as long as its trust value was higher than the threshold of 0.8.
- *T3U3*. In this condition, node D could behave untruthfully to three partner nodes of node A. It was identified that the trust value of node D could keep decreasing at most cases and fall below the threshold after 15 days without going up to the threshold again. Intuitively, the detection accuracy was better than for the first condition, as one more node would report a "malicious" status for node D.

The results demonstrate that our attack has the potential to compromise the robustness of the challenge mechanism in this scenario. If behaving normally to most partner nodes, a malicious node can maintain its trust value close to the threshold.

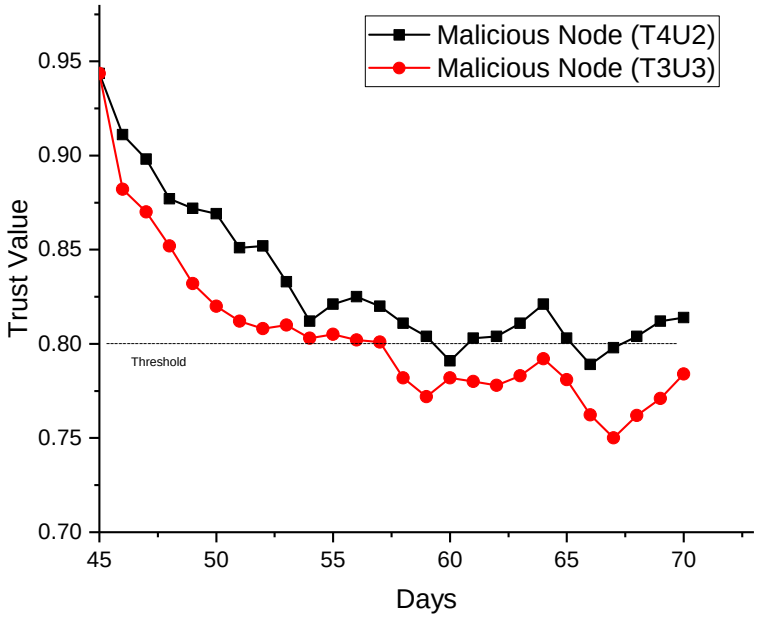

Figure 4. Trust values of malicious node (e.g., node D) calculated by target node (e.g., node A) under Scenario 1.

Results under Scenario 2

In this scenario, we randomly selected one expert node ($I = 0.95$) as a malicious node (e.g., node D), which conducted our special attack from day 45. Node D had to behave truthfully to node A, as node A could directly send challenges to node D. The trust computation of target node (node A) regarding node D is shown in Figure 5. We also tested two conditions: (1) node D behaving truthfully to four partner nodes of node A, whereas it provided untruthful feedback to the remaining two partner nodes (T4U2); (2) node D responding truthfully to three partner nodes but giving untruthful answers to the remaining three partner nodes (T3U3). It is worth noting that node D always sent truthful feedback to node A. The trust value of the malicious node is shown in Figure 5.

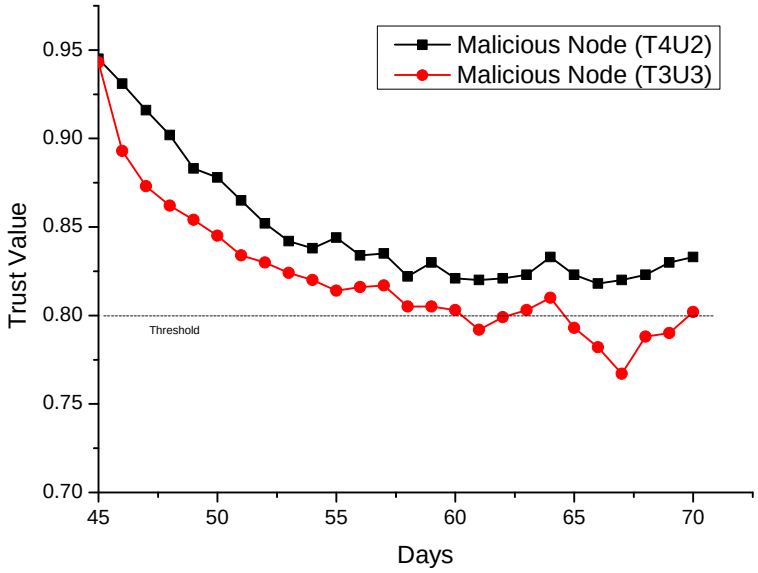

Figure 5. Trust values of malicious node (e.g., node D) calculated by target node (e.g., node A) under Scenario 2.

- *T4U2*. In this condition, the trust value of node *D* computed by node *A* could gradually decrease closer to the threshold during the first 10 days, because two partner nodes could report malicious actions of node *D* to node *A*. Afterwards, the trust value was maintained in the range from 0.81 to 0.82 in most cases, as four partner nodes reported that node *D* was normal. As the trust value was higher than the threshold of 0.8, node *D* could still make an influence on node *A* and its alarm aggregation.
- *T3U3*. In this condition, node *D* sent truthful feedback to three partner nodes of node *A* but sent malicious feedback to the other three partner nodes. It was found that the trust value of node *D* computed by node *A* could keep decreasing during the first 15 days, whereupon it was maintained around the threshold. As node *D* always sent truthful feedback to node *A*, its trust value crossed below and above the threshold.

As compared to the results reported in [6,7], the experimental results showed that our attack could greatly degrade the effectiveness and robustness of challenge-based CIDNs, for which a malicious node has a non-trivial chance to maintain its trust value above the threshold and affect the alarm aggregation for a target node. Further, we noticed that the trust value of a malicious node decreased faster in Scenario 1 than in Scenario 2. This was because node *D* could send truthful answers to node *A* in Scenario 2, as they were partner nodes, while in Scenario 1, node *A* could only evaluate the trustworthiness of node *D* on the basis of the feedback from its partner nodes.

5.3. Evaluation in a Real Environment

To explore the practical performance of the SOOA, we collaborated with an IT company and conducted a real evaluation in a wired CIDN including 26 nodes. Figure 6 shows the high-level network deployment. Each node could access the Internet by connecting to a server, which acted similarly to a firewall and provided various computing resources.

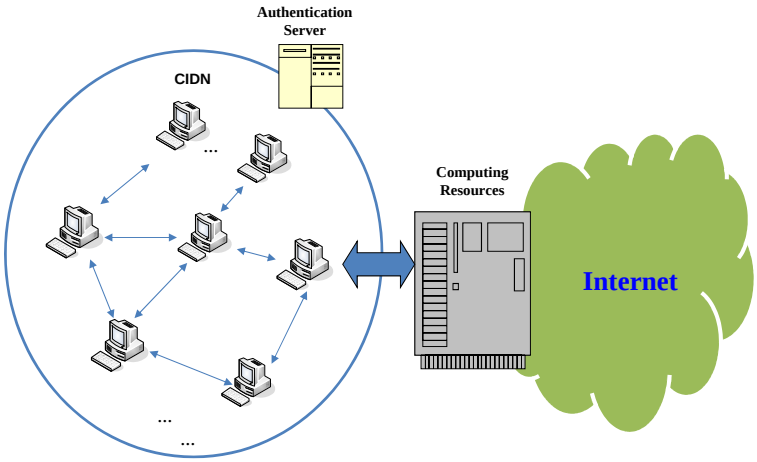

Figure 6. The high-level architecture of a real collaborative intrusion detection network (CIDN) environment.

To validate the results obtained in the simulated environment, we adopted the same environmental settings and observed the network to become stable; that is, the trust values turned out to be stable. We then set up the same scenarios and randomly selected one expert node to launch our attack. Figure 7 depicts the trust value of the malicious node under both Scenario 1 and Scenario 2.

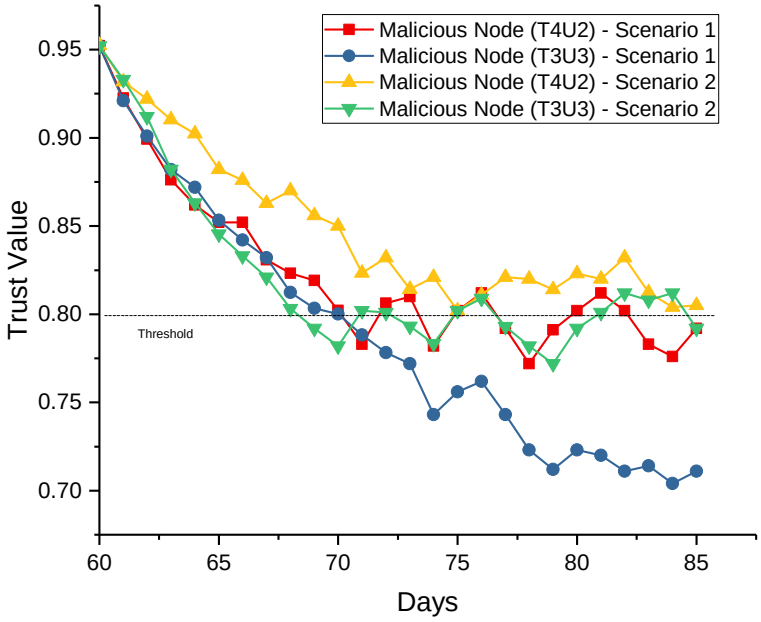

Figure 7. Trust value of malicious node under different scenarios in a real environment.

- *Scenario 1*. In this scenario, it was found that the trust value of the malicious node could decrease much faster under T3U3 than T4U2. More specifically, the trust value under T4U2 could deviate around the threshold, while decreasing nearly directly under T3U3. The observations are in-line with the results obtained in the simulated environment. While we identified that the trust value of the malicious node decreased faster in a real network environment under T3U3, an attacker should make a better strategy to launch the SOOA, that is, behave normally to more than half the nodes.
- *Scenario 2*. Figure 7 shows that the trust value of the malicious node under T3U3 could maintain a downtrend for the first 10 days but would deviate around the threshold later; that is, it was above the threshold for a total of 10 days. Under T4U2, it was found that the trust value of the malicious node would not decrease below the threshold. Similarly, the observations are in-line with the results obtained in the simulated evaluation.

On the whole, the experimental results in the real CIDN environment validated that our SOOA has the potential to compromise the robustness of the challenge mechanism by slowing the detection speed of malicious nodes. In Scenario 2, malicious nodes may have a greater chance to bypass the detection than in Scenario 1, as they can pretend to be normal to the target node directly. Furthermore, our attack could help to maintain the trust values of malicious nodes by employing a good strategy, for example, T4U2, behaving normally to most partners of the target node.

5.4. Discussion

The above experimental results have demonstrated the potential of the SOOA in compromising the robustness of challenge-based CIDNs, that is, slowing the detection speed of malicious nodes. As this is an initial study in this field, there are some improvements that can be made in our future work.

- *The impact of partner nodes' selection.* In this work, we assume that each partner has the same impact on the decision in the target node. Thus, there is no need to consider how to select a partner node for the SOOA. However, if we consider that the target node may give different weights to its partner nodes, then there is a need to identify which type of partner node can be attacked. This is an interesting topic for our future work.

- *A variety of combinations.* In this work, we only evaluated two combinations, namely, T3U3 and T4U2, under two scenarios. The obtained results have demonstrated the influence of the SOOA on the robustness of the challenge mechanisms. In future work, it will be an interesting topic to investigate the trend of trust values in other combinations, for example, T4U3 and T5U5.

To defend against this type of attack, there is a need to consider deploying additional security tools to enhance the performance of challenge mechanisms.

- *Emphasizing the impact of malicious actions.* In CIDNs, if a node wishes to evaluate a node's trustworthiness, it has to collect information from other trusted nodes. Our attack allows malicious nodes to behave maliciously, without a timely detection. One potential solution is to punish more for malicious actions, if detected by a node, for example, by building a trust computation by means of IS, which can give greater weight to expert nodes.
- *Employing additional measurements.* In challenge-based CIDNs, the trustworthiness of a node is mainly determined by challenges, but the challenges have to be sent over a period of time, rendering the network vulnerable to advanced attacks. To enhance the robustness of CIDNs, additional measures should be considered to evaluate the trustworthiness of a node, for example, packet-level trust [11].

Overall, our work validates that existing challenge mechanisms would suffer from advanced insider attacks in practice, as a result of the adopted assumptions. Additional security mechanisms can be considered to enhance the robustness of challenge-based CIDNs.

6. Conclusions

Challenge mechanisms are proved to be robust to common insider attacks, which evaluate the trustworthiness of others by sending challenges. However, we found that such kinds of mechanisms may still suffer from advanced insider attacks, as a result of the adopted assumptions. In this work, we develop a SOOA, which can behave normally to one or several nodes while sending untruthful answers to other nodes. To explore the attack performance, we performed two major experiments under both simulated and real CIDN environments. The experimental results indicate that our attack enables insider nodes to behave maliciously without being detected timely. Future work will include exploring how to enhance the existing framework to protect against advanced insider attacks, that is, applying the concept of IS.

Author Contributions: W. Li and W. Meng initialized the idea, performed the experiments and analyzed the data; L.F. Kwok designed the experiments and contributed to the paper writing.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IDS	Intrusion detection system
CIDN	Collaborative intrusion detection network
SOOA	Special on–off attack
IS	Intrusion sensitivity
PMFA	Passive message fingerprint attack

References

1. Scarfone, K.; Mell, P. *Guide to Intrusion Detection and Prevention Systems (IDPS)*; NIST Special Publication: Gaithersburg, MD, USA, 2007; pp. 800–894.
2. Gong, F. *Next Generation Intrusion Detection Systems (IDS)*; McAfee Network Security Technologies Group: Santa Clara, CA, USA, 2003.

3. Duma, C.; Karresand, M.; Shahmehri, N.; Caronni, G. A Trust-Aware, P2P-Based Overlay for Intrusion Detection. In Proceedings of the 17th International Workshop on Database and Expert Systems Applications, Krakow, Poland, 4–8 September 2006; pp. 692–697.
4. Wu, Y.-S.; Foo, B.; Mei, Y.; Bagchi, S. Collaborative Intrusion Detection System (CIDS): A Framework for Accurate and Efficient IDS. In Proceedings of the 2003 Annual Computer Security Applications Conference (ACSAC), Las Vegas, NV, USA, 8–12 December 2003; pp. 234–244.
5. Fung, C.J.; Boutaba, R. Design and management of collaborative intrusion detection networks. In Proceedings of the 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM), Ghent, Belgium, 27–31 May 2013; pp. 955–961.
6. Fung, C.J.; Baysal, O.; Zhang, J.; Aib, I.; Boutaba, R. Trust Management for Host-Based Collaborative Intrusion Detection. In *DSOM 2008, Lecture Notes in Computer Science (LNCS) 5273*; De Turck, F., Kellerer, W., Kormentzas, G., Eds.; Springer: Berlin, Germany, 2008; pp. 109–122.
7. Fung, C.J.; Zhang, J.; Aib, I.; Boutaba, R. Robust and scalable trust management for collaborative intrusion detection. In Proceedings of the 11th IFIP/IEEE International Conference on Symposium on Integrated Network Management (IM), Long Island, NY, USA, 1–5 June 2009; pp. 33–40.
8. Li, W.; Meng, Y.; Kwok, L.-F.; Ip, H.H.S. PMFA: Toward Passive Message Fingerprint Attacks on Challenge-based Collaborative Intrusion Detection Networks. In Proceedings of the 10th International Conference on Network and System Security (NSS), Taipei, Taiwan, 28–30 September 2016; pp. 433–449.
9. Meng, M.; Luo, X.; Li, W.; Li, Y. Design and Evaluation of Advanced Collusion Attacks on Collaborative Intrusion Detection Networks in Practice. In Proceedings of the 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Tianjin, China, 23–26 August 2016; pp. 1061–1068.
10. Li, W.; Meng, Y.; Kwok, L.-F. SOOA: Exploring Special On-Off Attacks on Challenge-Based Collaborative Intrusion Detection Networks. In Proceedings of the 12th International Conference on Green, Pervasive, and Cloud Computing (GPC), Cetara, Italy, 11–14 May 2017; pp. 402–415.
11. Meng, Y.; Li, W.; Xiang, Y.; Choo, K.-K.R. A Bayesian Inference-based Detection Mechanism to Defend Medical Smartphone Networks against Insider Attacks. *J. Netw. Comput. Appl.* **2017**, *78*, 162–169.
12. Donovan, S.; Feamster, N. Alternative trust sources: Reducing DNSSEC signature verification operations with TLS. In Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication (SIGCOMM), London, UK, 17–21 August 2015; pp. 353–354.
13. Kolias, C.; Kolias, V.; Kambourakis, G. TermID: A distributed swarm intelligence-based approach for wireless intrusion detection. *Int. J. Inf. Secur.* **2017**, *16*, 401–416.
14. Meng, Y.; Li, W.; Kwok, L.-F. Evaluation of detecting malicious nodes using Bayesian model in wireless intrusion detection. In Proceedings of the 7th International Conference on Network and System Security (NSS), Helsinki, Finland, 21–23 August 2013; Lecture Notes in Computer Science 7873; Springer: Berlin, Germany, 2013; pp. 40–53.
15. Meng, W.; Au, M.H. Towards statistical trust computation for medical smartphone networks based on behavioral profiling. In Proceedings of the 11th IFIP WG 11.11 International Conference on Trust Management (IFIPTM), Gothenburg, Sweden, 12–16 June 2017; pp. 152–159.
16. Meng, W.; Li, W.; Wang, Y.; Au, M.H. Detecting malicious nodes in medical smartphone networks through euclidean distance-based behavioral profiling. In Proceedings of the 9th International Symposium on Cyberspace Safety and Security (CSS), Xi'an, China, 23–25 October 2017; pp. 163–175.
17. Meng, W.; Li, W.; Su, C.; Zhou, J.; Lu, R. Enhancing Trust Management for Wireless Intrusion Detection via Traffic Sampling in the Era of Big Data. *IEEE Access* **2017**, doi:10.1109/ACCESS.2017.2772294.
18. Ramos, A.; Lazar, M.; Filho, R.H.; Rodrigues, J.J.P.C. A security metric for the evaluation of collaborative intrusion detection systems in wireless sensor networks. In Proceedings of the IEEE International Conference on Communications (ICC), Paris, France, 21–25 May 2017.
19. Li, W.; Meng, Y.; Kwok, L.-F.; Ip, H.H.S. Developing Advanced Fingerprint Attacks on Challenge-based Collaborative Intrusion Detection Networks. *Clust. Comput.* **2017**, 1–12, doi:10.1007/s10586-017-0955-8.
20. Li, W.; Meng, Y.; Kwok, L.-F. Enhancing Trust Evaluation Using Intrusion Sensitivity in Collaborative Intrusion Detection Networks: Feasibility and Challenges. In Proceedings of the 9th International Conference on Computational Intelligence and Security (CIS), Leshan, China, 14–15 December 2013; pp. 518–522.

21. Li, W.; Meng, W.; Kwok, L.-F. Design of intrusion sensitivity-based trust management model for collaborative intrusion detection networks. In *Trust Management VIII, IFIP AICT*; Zhou, J., Gal-Oz, N., Zhang, J., Gudes, E., Eds.; Springer: Heidelberg, Germany, 2014; Volume 430, pp. 61–76.
22. Li, W.; Meng, Y.; Kwok, L.-F.; Ip, H.H.S. Enhancing Collaborative Intrusion Detection Networks Against Insider Attacks Using Supervised Intrusion Sensitivity-Based Trust Management Model. *J. Netw. Comput. Appl.* **2017**, *77*, 135–145.
23. Li, W.; Meng, W. Enhancing collaborative intrusion detection networks using intrusion sensitivity in detecting pollution attacks. *Inf. Comput. Secur.* **2016**, *24*, 265–276.
24. Cho, J.-H.; Chan, K.; Adali, S. A Survey on Trust Modeling. *ACM Comput. Surv.* **2015**, *48*, doi:10.1145/2815595.
25. Douceur, J. The sybil attack. In *IPTPS 2002. LNCS*; Druschel, P., Kaashoek, M.F., Rowstron, A., Eds.; Springer: Heidelberg, Germany, 2002; Volume 2429.
26. Perrone, L.P.; Nelson, S.C. A Study of On-Off Attack Models for Wireless Ad Hoc Networks. In Proceedings of the 2006 Workshop on Operator-Assisted Community Networks, Berlin, Germany, 18–19 September 2006; pp. 1–10.
27. Snort: An Open Source Network Intrusion Prevention and Detection System (IDS/IPS). Homepage. Available online: http://www.snort.org/ (accessed on 8 January 2018).